TITLE:

Protein Structure Prediction by Evaluating Sequence-Structure Compatibility

AUTHOR(S):

Matsuo, Yo

CITATION:

Matsuo, Yo. Protein Structure Prediction by Evaluating Sequence-Structure Compatibility. Bulletin of the Institute for Chemical Research, Kyoto University 1994, 72(3-4): 319-329

ISSUE DATE:

1994-11-30

URL:

http://hdl.handle.net/2433/77585

RIGHT:

京都大学学術情報リポジトリ
KURENAI 紅
Kyoto University Research Information Repository

Bull. Inst. Chem. Res., Kyoto Univ., Vol. 72, No. 3–4, 1994

Protein Structure Prediction by Evaluating Sequence-Structure Compatibility

Yo Matsuo*

Received June 15, 1994

A method for protein structure prediction has been developed. The method evaluates the compatibility of a given sequence with known structures in terms of side-chain packing, solvation, hydrogen-bonding and local conformation, and identifies the most likely structure. The method has been applied to a large number of proteins. Here, structure predictions for the following proteins are described in detail; spermidine/putrescine-binding protein, shikimate kinase, and the hydrophilic subunit of mannose permease. The predictions suggest possible residues of functional importance as well as evolutionary relationships with other proteins.

KEY WORDS: Structure prediction/ Sequence-structure compatibility/ Evolutionary relationship

1. INTRODUCTION

The number of folds adopted by proteins is believed to be limited, probably as low as 1,000.[1] There are in fact many examples of proteins which, despite no significant sequence similarity, adopt similar structures (e.g., actin and 70 kD heat shock protein[2]). From these observations, several authors have developed methods for protein structure prediction, which evaluate the compatibility of a sequence with known structures.[3-8] Using these methods, one can frequently identify the most likely structure of a protein from a library of known structures.

We have developed our own method.[9] It uses four functions: side-chain packing, solvation, hydrogen-bonding and local structure functions. Those functions are combined to give a score which measures the sequence-structure compatibility. The method has been applied to a large number of proteins in protein sequence databases. In the present paper, structure predictions for the following proteins are described in detail; spermidine/putrescine-binding protein, shikimate kinase, and the hydrophilic subunit of mannose permease. The functional and evolutionary implications of the predictions are discussed.

2. MATERIALS AND METHODS

2.1 *Evaluation of sequence-structure compatibility*

Four functions were used for the evaluation of the sequence-structure compatibility: side-chain packing (F_{sp}), solvation (F_{solv}), hydrogen-bonding (F_{hb}), and local structure (F_{loc}) functions. Except for F_{sp}, they were defined in the same way as in our previous work.[9] They have the following general form:

* 松尾　洋: Protein Engineering Research Institute, 6-2-3 Furuedai, Suita, Osaka 565, Japan.

$$Fx(a\,;\ s)=-\log\,(fx(s)),\ \ (x=\{solv,\ hb,\ loc\}),$$

where a denotes a type of amino acid (for F_{solv} and F_{loc}) or amino acid pair (for F_{hb}) ; s, the state of a (solvent-accessibility for F_{solv}, hydrogen-bonded or not for F_{hb}, and local structure for F_{loc}) ; $f_x(a\,;\ s)$, the frequency of a in the state s ; and $f_x(s)$, the frequency of any amino acid or amino acid pair in the state s.

The side-chain packing function (F_{sp}) has been improved to take into account inter-residue contact and angle as well as distance.[10,11] F_{sp} indicates the propensity of the amino acid pair (a, b) to be in contact in a particular spatial relationship. The spatial relationship between two residues was defined by the distance (d) between the Cβ atoms of the residues and the angle (θ) between the residues. The angle θ between residues i and j was defined as the sum of the angles Cβ_i-Cα_i-Cβ_j and Cβ_j-Cα_j-Cβ_i. Here, Cα_i and Cβ_i denote the Cα and Cβ atoms of the residue i. For glycines, virtual Cβ atoms were generated according to a standard amino acid conformation. Then, F_{sp} was defined by:

$$F_{sp}(a, b\,;\ d,\ \theta)=w(a, b\,;\ d,\ \theta)\,\{dE_0(a, b)+dE((a, b\,;\ d,\ \theta)\},$$

Here,

$$w(a, b\,;\ d,\ \theta)=NC(a, b\,;\ d,\ \theta)/N(a, b\,;\ d,\ \theta),$$

$$dE_0(a, b)=-\log\left\{\frac{NC(a, b)/N10(a, b)}{NC/N10}\right\},$$

$$dE(a, b\,;\ d,\ \theta)=-\log\left\{\frac{NC(a, b\,;\ d,\ \theta)/NC(a, b)}{NC(d,\ \theta)/NC}\right\}.$$

$NC\,(a, b\,;\ d,\ \theta)$ is the number of observations of the residue pair (a, b) being in contact at distance d and angle θ; $N\,(a, b\,;\ d,\ \theta)$, that of (a, b) being at d and θ; $NC\,(a, b)$, that of (a, b) being in contact; $N10\ (a\ b)$, that of (a, b) being within 10 Å from each other ; NC , that of any residue pair being in contact; $N10$, that of any residue pair being within 10 Å from each other ; $NC(d,\ \theta)$, that of any residue pair being in contact at distance d and angle θ. Now, $dE_0(a, b)$ describes the tendency of (a, b) to be in contact, and $dE(a, b\,;\ d,\ \theta)$ is the preference of (a, b) for being at d and θ on the condition that they are in contact.

The details of the definitions of the functions are described elsewhere.[12] The parameters defining the functions were derived using the set of the coordinate data of 101 proteins taken from Protein Data Bank.[13] They have less than 30% sequence identity with one another and their resolutions are better than 2.5 Å.

A sequence was aligned with a structure using the Needleman-Wunsch algorithm[14] and a residue position dependent scoring table.[4] For a structure, a scoring table was constructed with the frozen approximation ;[6,7] the compatibility score for an amino acid at a residue position of a structure was calculated keeping the native residues of the structure at the remaining positions.

For a sequence aligned with a structure, scores S_x ($x=\{sp,\ solv,\ hb,\ loc\}$) were given by summing up the values of F_x over all residues (or residue pairs) of the sequence. S_x were then summed up to give S_{tot}, which measured the compatibility of the sequence with the structure. A negatively large score indicates better compatibility.

The performance of the functions was tested following the procedure of Hendlich *et al.*.[15] The test was to identify the native structure of a given protein from a large number of incorrect structures of the same length. The test was done with the Jack-knife procedure, where a test

protein was removed from the data set for the parameter calculation. Of the 100 proteins tested, 99 were correctly identified by the S_{tot} scores.[12] Cytochrome c_3 was the only exception. It has four haems. Such prosthetic groups were ignored in this work. This might be why cytochrome c_3 was not identified.

2.2 *Compatibility search*

A given protein sequence was compared with a library of known structures using the above procedure. The structures in the library were selected from PDB. They showed less than 30% sequence identity with one another. For the individual structures, compatibility scores S_{tot} were calculated. The scores were then normalized using the mean and standard deviation over all the structures. The normalization facilitated the evaluation of the statistical significance of the scores. It was empirically found that a score of -3.0 or better indicates good sequence-structure compatibility. The best scored structure was considered as the most likely structure for the sequence.

2.3 *Sequence comparison*

The Needleman-Wunsch method[14] and the PAM250 matrix[16] were used for comparing amino acid sequences. Statistical significance of the sequence similarity was evaluated by a jumbling test with 100 pairs of randomized sequences. A multiple sequence alignment was made with a pairwise based method[17] using the minimum spanning tree algorithm.

2.4 *Sequence motif search*

The NBRF-PIR sequence database (sections 1 and 2 of release 38, 30 Sep 1993; 43,658 sequences, 13,021,641 residues) was searched for proteins with a query sequence motif. In the search, small conservative substitutions were allowed. For all subsequences of the same length, similarity scores were measured by the PAM250 matrix.[16] If a subsequence showed more than 85% of the score for the exact match with the motif, then it was retained.

2.5 *Structure modeling*

A structure model of spermidine/putrescine-binding protein was built using the structure of maltose-binding protein as a template (see RESULTS AND DISCUSSION). The proteins were aligned as described above. Around the regions where deletions/insertions of residues occurred, backbone structures were constructed with the loop search method of Jones and Thirup.[18] The program FRGMNT[19] was used for the loop search. The side-chain conformations were modeled with the dead-end elimination algorithm.[20,21] Energy minimization was done using PRESTO.[22]

3. RESULTS AND DISCUSSION

3.1 *Spermidine/putrescine-binding protein*

Spermidine/putrescine-binding protein (SPBP; 39 kDa, 348 amino acids) is the periplasmic component of the spermidine/putrescine transport system of *E. coli*.[23] Its amino acid sequence was compared with a library of 131 known structures including the following other periplasmic binding proteins: maltose-(MBP), sulfate-(SBP), arabinose-(ABP), galactose/glucose-(GGBP), ribose-(RBP), and leucine-(LBP) binding proteins. Although these binding proteins show a variety of ligand specificities and lack significant sequence similarity, they are known to share a

similar fold.

Among the structures in the library, MBP (43 kDa, 396 amino acids) showed an extremely low compatibility score (−3.66), and this suggested that SPBP may adopt a similar structure (Table 1). Of the periplasmic binding proteins, SBP showed the second best score (−1.87).

Table 1. Compatibility of SPBP sequence with known structures; 131 structures were compared and sorted in order of their compatibility scores. The best 10 structures are listed below.

Rank	Structure	PDB code	Compatibility score
1	Maltose-binding protein	1OMP	−3.66
2	*p*-Hydroxybenzoate hydroxylase	1PHH	−2.01
3	Isocitrate dehydrogenase	3ICD	−1.97
4	Sulfate-binding protein	1SBP	−1.87
5	Actin	1ATN (A)	−1.83
6	Ribose-binding protein	1DRI	−1.57
7	Galactose/glucose-binding protein	2GBP	−1.55
8	Phosphofructokinase	1PFK (A)	−1.53
9	Malate dehydrogenase	4MDH (A)	−1.50
10	Leucine-binding protein	2LBP	−1.50

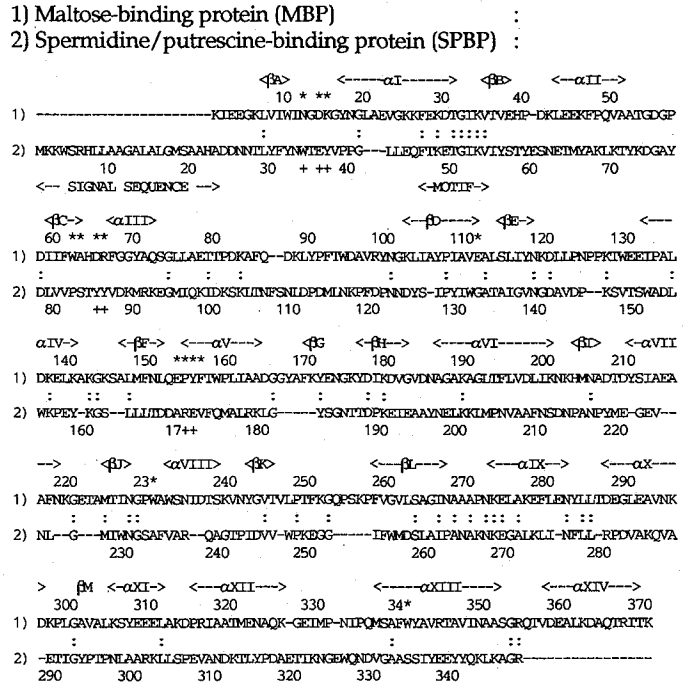

Fig. 1. Alignment of SPBP sequence with MBP structure. ':' indicates residue conservation; '*', ligand-binding residues of MBP; and '+', residues of SPBP which might be involved in ligand binding. Residue numbering for MBP here is different from Table 2, where the 26-residue signal sequence is included.

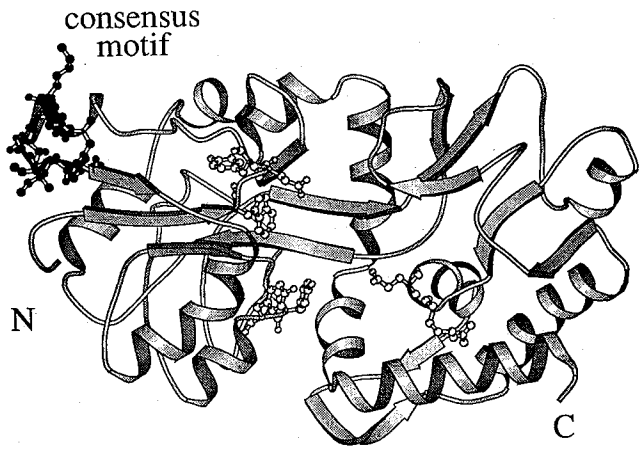

Fig. 2. A structure model of SPBP. Residues of the consensus motif are noted by black balls. Those which might be involved in the substrate binding are noted by white balls. The drawing was produced with MOLSCRIPT.[36]

This is consistent with the observation [24] that SBP is more similar in structure to MBP than ABP, GGBP, RBP and LBP.

SPBP sequence was aligned with MBP structure (Fig. 1). A model of SPBP structure was built using the alignment and the MBP coordinate data[24] (Fig. 2). The strands G, H and helix XIV, which form a small domain in MBP,[24] are missing in SPBP. In MBP, hydrogen-bonds and van der Waals contacts with maltose are mainly formed by charged or aromatic residues from the loops located in the cleft between the two domains.[24] According to the alignment, the following residues of SPBP might be involved in the ligand binding : Trp34, Glu36, Tyr37, Tyr85, Tyr86, Arg170 and Glu171.

The alignment revealed a highly conserved sequence motif in the loop region between the first α-helix (helix I) and the second β-strand (strand B). The motif spans residues 53 to 61 ('FEKDTGIKV') of MBP and residues 46 to 54 ('FTKETGIKV') of SPBP. The PIR sequence database was searched for similar sequence patterns. Out of 43,658 sequences in the database,

Table 2. Proteins which have sequence patterns similar to 'FTKETGIKV' or 'FEKDTGIKV'.

PIR code	Protein	Residues	Sequence
D40840	*E. coli* SPBP	46～54	FTKETGIKV
JGECM	*E. coli* MBP	53～61	FEKDTGIKV
S05330	*Enterobacter aerogenes* MBP	53～61	FEKDTGIKV
S05331	*Salmonella typhimurium* MBP	53～61	FEQDTGIKV
QRSEUA	*Serratia marcescens* IBP	51～59	FTKDTGIKV
B60816	*Neisseria meningitidis* IBP	42～50	FTRATGIKV
S10256	*Neisseria gonorrhoeae* IBP	42～50	FTRATGIKV
S26445	*Methanobacterium thermoformicicum* plasmid pFV1 methyltransferase EcoRII	40～48	FEKNHGIKI

only 8 were found to have similar patterns (Table 2): SPBP from *E. coli*; MBPs from *E. coli*, *Enterobacter aerogenes* and *Salmonella typhimurium*; IBPs from *Serratia marcescens*, *Neisseria meningitidis* and *Neisseria gonorrhoeae*; and methyltransferase EcoRII from *Methanobacterium thermoformicicum* plasmid pFV1. Except for methyltransferase EcoRII, which might represent noise in the database search, all the proteins were periplasmic binding proteins of Gram-negative bacteria. They covered all SPBP, MBP and IBP sequences in the database. The sequence similarity of IBPs to MBPs and SPBP is also a new finding. From the sequence alignment, a consensus pattern of 'F(T/E)(K/R/Q)(D/E/A)TGIKV' was observed, where (T/E) denotes T or E, and so on. The high specificity of the motif to the three periplasmic binding proteins SPBP, MBP and IBP suggests a common functional role of the motif in the transport systems. The motif is located on the surface loop of the N-terminal domain, which is apart from the ligand-binding cleft (Fig. 2). The motif might be involved in the interactions with the membrane components of the transport system, rather than the ligand binding.

Table 3. Sequence similarity among periplasmic binding proteins; SPBP *(E. coli)*, IBPs from *Serratia marcescens* (sIBP) and *Neisseria meningitidis* (nIBP), MBP *(E. coli)*, SBP *(Salmonella typhimurium)*, RBP *(E. coli)*, ABP *(E. coli)*, GGBP *(E. coli)*, LBP *(E. coli)*, and phosphate-binding protein (PBP) from *E. coli*. The upper right triangle shows the significance of the sequence similarity in units of standard deviations above the mean derived from jumbling tests. The lower left triangle shows % sequence identity.

	SPBP	sIBP	nIBP	MBP	SBP	RBP	ABP	GGBP	LBP	PBP
SPBP		4.35	4.49	2.38	3.49	0.82	0.84	0.17	−0.56	0.94
sIBP	22.5		25.89	4.87	0.29	0.75	0.42	−0.48	0.82	1.38
nIBP	21.8	37.9		4.71	2.57	1.37	1.08	1.56	0.73	0.68
MBP	22.1	23.1	25.8		1.72	0.75	1.61	0.19	1.22	1.37
SBP	20.7	20.4	21.7	21.0		0.73	1.25	2.82	0.81	1.48
RBP	22.1	24.4	21.0	23.2	18.1		8.30	10.50	2.02	0.90
ABP	19.7	17.7	17.0	18.7	19.7	24.4		5.83	1.09	1.60
GGBP	19.7	17.2	18.8	19.4	16.5	26.9	22.0		1.36	−1.99
LBP	14.2	16.3	19.4	20.5	18.4	22.5	17.0	19.7		1.19
PBP	16.8	20.6	19.9	21.5	18.4	21.8	15.1	15.2	18.4	

The periplasmic binding proteins show little sequence similarity with one another. However, the jumbling test detected statistical significance of 4 standard deviation units or more among SPBP, two IBPs and MBP (Table 3). This is consistent with the structural analysis by Spurlino *et al.*,[24] where MBP and SBP are classified into a different group from the other binding proteins. Together with the sequence-structure compatibility and the existence of the conserved sequence motif, a weak, but overall sequence similarity suggests that SPBP, MBP and IBP have an evolutionary relationship.

X-ray analysis of SPBP is now in progress (S. Sugiyama and K. Morikawa, personal communication). Their results will enable us to assess the efficacy of our method.

3.2 *Shikimate kinase*

Shikimate kinase (SKase; EC 2.7.1.71; 19 kDa) catalyzes the phosphorylation of shikimic acid to shikimate 3-phosphate in the shikimate pathway for aromatic amino acid biosynthesis in plants and microorganisms. SKase sequences in the PIR database were compared with known

Table 4. Similarity between SKase sequences and AKase sequence, and their compatibility with AKase structure. The upper right triangle shows the significance of the sequence similarity in units of standard deviations above the mean derived from jumbling tests. The lower left triangle shows % sequence identity. The right most column shows the sequence-structure compatibility scores. SKII, *E. coli* SKase II; SK, *Erwinia chrysanthemi* shikimate kinase; ARO1, yeast ARO1 protein; aroM, *Emericella nidulans* aroM protein; AK, pig adenylate kinase.

	SKII	SK	ARO1	aroM	AK	Score
SKII		30.15	8.21	10.16	5.13	−3.40
SK	51.4		11.23	10.05	4.84	−3.22
ARO1	27.0	28.9		16.10	3.00	−2.53
aroM	28.2	30.1	39.7		1.91	−1.74
AK	19.0	19.7	16.1	17.1		−5.08

structures. Pig adenylate kinase (AKase; EC 2.7.4.3; 22 kDa, 194 amino acids) structure (PDB code, 3ADK) showed good compatibility scores (Table 4). This suggests the structural similarity between SKase and AKase.

The type A sequence motif 'GXXXXGK(S/T)' is found in various ATP- or GTP-binding proteins.[25] The alignment of AKase structure and SKase sequences (Fig. 3) shows that they have the motif in the similar regions of the sequences (Gly15 to Gly22 of AKase; Gly9 to

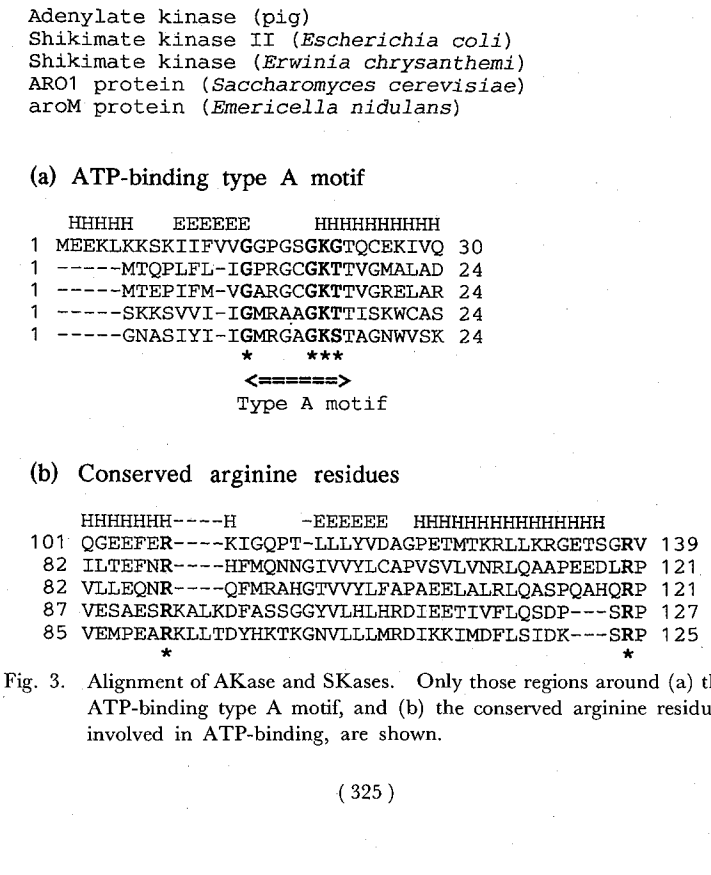

Fig. 3. Alignment of AKase and SKases. Only those regions around (a) the ATP-binding type A motif, and (b) the conserved arginine residues involved in ATP-binding, are shown.

Thr/Ser16 of SKases). The alignment also revealed two conserved arginine residues : Arg107 and 138 (Arg88 and 120) of AKase (*E. coli* SKase II) (Fig. 3). The Arg107 of AKase is a part of the type B sequence motif for ATP binding.[25] The Arg138 is also known to be involved in ATP binding.[26] The conservation of these functionally important residues supports the structure prediction, and suggests the similarity between the functional mechanisms of AKase and SKase.

Although overall sequence similarity between AKase and SKases is very low (less than 20% identity; Table 4), the jumbling test showed weak statistical significance of similarity of *E. coli* SKase II and *Erwinia chrysanthemi* SKase to pig AKase by 5.13 and 4.84 in standard deviation unit, respectively. This weak sequence similarity might indicate distant evolutionary relationship, together with the sequence-structure compatibility and the conservation of functionally important residues.

3.3 *Hydrophilic subunit of mannose permease*

The mannose permease of *E. coli* is a component of the phosphotransferase system (PTS). It mediates the transport of mannose and related hexoses across the cytoplasmic membrane. The permease consists of a hydrophilic subunit $IIAB^{Man}$, and two transmembrane subunits IIC^{Man} and IID^{Man}. $IIAB^{Man}$ (35 kDa, 323 amino acids) catalyzes the phosphate transfer from histidine-containing phosphocarrier protein (HPr) to the sugar substrate.

Table 5. Compatibility of $IIAB^{Man}$ sequence with known structures. The structures were sorted in order of their compatibility scores. The best 10 structures are listed below.

Rank	Structure	PDB code	Compatibility score
1	Galactose/glucose-binding protein	2GBP	−3.34
2	Leucine-binding protein	2LBP	−2.53
3	Malate dehydrogenase	4MDHA	−2.24
4	Arabinose-binding protein	8ABP	−1.80
5	D-Xylose isomerase	6XIA	−1.79
6	Lactate dehydrogenase	6LDH	−1.35
7	Tryptophan synthase β subunit	1WSYB	−1.34
8	Aspartate aminotransferase	2AAT	−1.30
9	Citrate synthase	2CTS	−1.27
10	Aconitase	5ACN	−1.25

$IIAB^{Man}$ sequence was compared with known structures. *E. coli* galactose/glucose binding protein (GGBP; 33 kDa, 309 amino acids) showed a very low compatibility score (−3.34). Other periplasmic binding proteins LBP and ABP also showed good scores (−2.53 and −1.80, respectively) (Table 5). It has been reported that $IIAB^{Man}$ consists of two structural domains IIA (14 kDa, residues 1～136) and IIB (20 kDa, residues 156～323), which are linked by an Ala-Pro-rich flexible hinge of 20 residues.[27] The IIA and IIB domains were aligned well with the N- and C-terminal domains of GGBP, respectively (Fig. 4). The Ala-Pro-rich hinge was aligned with the first α-helix of the C-terminal domain of GGBP (Fig. 4). The helix is a part of the hinge region which is important for inter-domain motion of the periplasmic binding proteins.[28]

$IIAB^{Man}$ is phosphorylated at His10 of IIA domain and His175 of IIB domain.[27] A phospho group is first transferred from HPr to His10, and next from His10 to His175, and finally

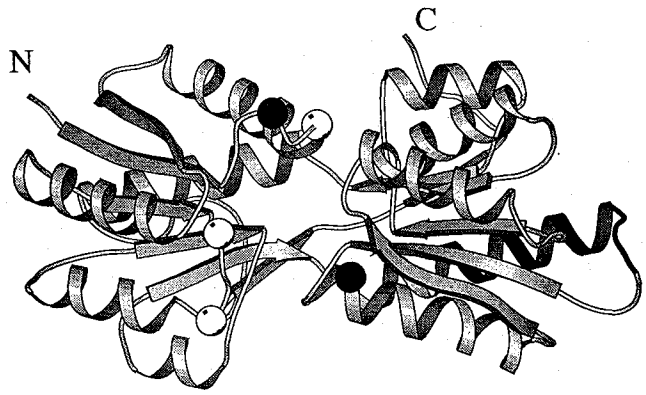

Fig. 4. $IIAB^{Man}$ sequence is threaded onto GGBP structure. Black balls indicate the phosphorylation residues His10 (upper left) and His175 (lower right); white balls, Trp12, Trp69 and Ser72; and black ribbon, the Ala-Pro-rich hinge. The drawing was produced with MOLSCRIPT.[36]

from His 175 to the sugar substrate. For this phosphotransfer to happen, the two histidine residues and the substrate binding region must be in close proximity. Our prediction satisfies this requirement; when $IIAB^{Man}$ sequence is threaded onto GGBP structure, His10 and His175 are located on loops in the substrate-binding cleft of GGBP (Fig. 4).

Recently, Markovic-Housley *et al.*[29] predicted that the IIA domain would be structurally similar to flavodoxin. Their prediction was based on experimental data and 3D profile methods.[4,30] First, mutant studies suggested that the residues Trp12, Trp69 and Ser72 are spatially close to the active site His10. Second, NMR studies indicated an alternating β/α structure consisting of 4 α-helices and 5 β-strands. Our prediction is consistent with these experimental data. Trp12, Trp69 and Ser72 are located on loops in the substrate-binding cleft, in close proximity of His10 (Fig. 4). The N-terminal domain of GGBP, with which IIA domain was aligned by our method, has the same topology with flavodoxin; the order of the five β-strands is 54312.

3.4 *Others*

In addition to those described above, a number of proteins which are likely to be compatible with known structures have been found[31] through the application of the present method to 11,706 sequences in PIR database release 38, section 1. Two of them are briefly described below.

First, it has been predicted that rat tyrosine aminotransferase (EC 2.6.1.5; 51 kDa, 454 amino acids) would be structurally similar to *E. coli* aspartate aminotransferase (EC 2.6.1.1; 44 kDa, 396 amino acids) (score -3.18). Although the two amino transferases lack significant overall sequence similarity (17% identity), they both catalyze the transfer of an α-amino group from an α-amino acid to an α-ketoglutarate, with pyridoxal 5′-phosphate (PLP) as a coenzyme. And several functionally important residues are conserved. From these observations, distant homology between the two aminotransferases was previously predicted by Hargrove *et al.*[32] and Mehta *et al.*.[33] Our prediction supports the previous predictions.

Second, it has been predicted that *E. coli* threonine dehydratase (TDH; EC 4.2.1.16; 35

kDa, 329 amino acids), *Corynebacterium glutamicum* threonine synthase (TSY ; EC 4.2.99.2 ; 37 kDa, 352 amino acids), and *E. coli* and *Salmonella typhimurium* cysteine synthase A (CSYA ; EC 4.2.99.8 ; 35 kDa, 323 amino acids) would be structurally similar to *E. coli* tryptophan synthase β subunit (WSYβ; EC 4.2.1.20; 43 kDa, 396 amino acids) (scores -3.06, -3.44, -3.88 and -3.77, respectively). TDH, TSY and CSYA have significant sequence homology with one another, and belong to the same superfamily (called TDH superfamily here). Although WSYβ lacks significant overall sequence similarity with TDH superfamily, they catalyze similar reactions with PLP as a cofactor and act on successive steps in metabolic pathways. And some functionally important residues are conserved among them. From these observations, distant homology between WSYβ and TDH superfamily was predicted by Levy and Danchin,[34] and Bork and Rohde.[35] Our results support their predictions.

3.5 *Future directions*

There are many examples of proteins which share similar local structures (domains, supersecondary structures, etc.) although their overall structures are different.[17] For example, the NAD-binding fold (Rossman fold) is commonly found in various dehydrogenases ; similar α/β folds are found in various ATP- or GTP-binding proteins, such as adenylate kinase, EF-Tu, rec A, ras p21 protein, etc. By adding such recurrent local structures to the library, the present method could have a wider application.

4. CONCLUSIONS

A method for protein structure prediction has been developed, which evaluates the compatibility of a sequence with known structures and identifies the most likely structure. Using the method, the structural similarity between SPBP and MBP, SKase and AKase, and $IIAB^{Man}$ and GGBP was predicted. The predictions suggested functionally important residues as well as evolutionary relationships with other proteins. The predictions would be useful for planning experiments, such as site-directed mutagenesis. Altogether, the present work demonstrated that the sequence-structure compatibility approach to structure prediction is very promising.

ACKNOWLEDGEMENTS

I am grateful to Ken Nishikawa for advice and encouragement. I thank Shigeru Sugiyama and Kosuke Morikawa for discussions on the prediction of spermidine/putrescine-binding protein structure, and Haruki Nakamura for discussions and advice on the structure modeling.

REFERENCES

(1) C. Chothia, *Nature,* **357**, 543 (1992).
(2) W. Kabsch, H.G. Mannherz, D. Suck, E.F. Pai, and K.C. Holmes, *Nature,* **347**, 37 (1990).
(3) M.J. Sippl, *J. Mol. Biol.,* **213**, 859 (1990).
(4) J.U. Bowie, R. Luthy and D. Eisenberg, *Science,* **253**, 164 (1991).
(5) D.T. Jones, W.R. Taylor and J.M. Thornton, *Nature,* **358**, 86 (1992).
(6) A. Godzik, A. Kolinski and J. Skolnick, *J. Mol. Biol.,* **227** (1992).
(7) M. Wilmanns and D. Eisenberg, *Proc. Natl. Acad. Sci. USA,* **90**, 1379 (1993).
(8) S.J. Wodak and M.J. Rooman, *Curr. Opin. Struct. Biol.,* **3**, 247 (1993).
(9) K. Nishikawa and Y. Matsuo, *Protein Eng.,* **6**, 811 (1993).

(10) Y. Matsuo and K. Nishikawa, *Protein Eng.*, **6**, 1027 (1993).
(11) Y. Matsuo and K. Nishikawa, *FEBS Lett.*, **345**, 23 (1994).
(12) Y. Matsuo, H. Nakamura and K. Nishikawa, submitted.
(13) F.C. Bernstein, T.F. Koetzle, G.T.B. Williams, E.F. Meyer, M.D. Brice, J.R. Rodgers, O. Kennard, T. Shimanouchi and M. Tasumi, *J. Mol. Biol.*, **112**, 535 (1977).
(14) S.B. Needleman and C.D. Wunsch, *J. Mol. Biol.*, **48**, 443 (1970).
(15) M. Hendlich, P. Lackner, S. Weitckus, H. Floeckner, R. Froschauer, K. Gottsbacher, G. Casari and M.J. Sippl, *J. Mol. Biol.*, **216**, 167 (1990).
(16) M.O. Dayhoff, R.M. Schwartz and B.C. Orcutt, in: Atlas of Protein Sequence and Structure, vol. 5, suppl. 3, 345 (1978).
(17) Y. Matsuo and M. Kanehisa, *CABIOS*, **9**, 153 (1993).
(18) T.A. Jones and S. Thirup, *EMBO J.*, **5**, 819 (1986).
(19) H. Nakamura, K. Katayanagi, K. Morikawa and M. Ikehara, *Nucl. Acids Res.*, **19**, 1817 (1991).
(20) J. Desmet, M. De Mayer, B. Hazes and I. Lasters, *Nature*, **356**, 539 (1992).
(21) R. Tanimura, A. Kidera, I. Fujii and H. Nakamura, *Protein Eng.*, **6**, 1023 (1993).
(22) K. Morikami, T. Nakai, A. Kidera, M. Saito and H. Nakamura, *Computers Chem.*, **16**, 243 (1992).
(23) T. Furuchi, K. Kashiwagi, H. Kobayashi and K. Igarashi, *J. Biol. Chem.*, **266**, 20928 (1991).
(24) J.C. Spurlino, G-Y. Lu and F.A. Quiocho, *J. Biol. Chem.*, **266**, 5202 (1991).
(25) J.E. Walker, M. Saraste, M.J. Runswick and N.J. Gay, *EMBO J.*, **1**, 945 (1982).
(26) D. Dreusicke, P.A. Karplus and G.E. Schulz, *J. Mol. Biol.*, **199**, 359 (1988).
(27) B. Erni, B. Zanolari, P. Graff and H.P. Kocher, *J. Biol. Chem.*, **264**, 18733 (1989).
(28) S.L. Mowbray and L.B. Cole, *J. Mol. Biol.*, **225**, 155 (1992).
(29) Z. Markovic-Housley, J. Balbach, B. Stolz and J-C. Genovesio-Taverne, *FEBS Lett.*, **340**, 202 (1994).
(30) M.S. Johnson, J.P. Overington and T.L. Blundell, *J. Mol. Biol.*, **231**, 735 (1993).
(31) Y. Matsuo and K. Nishikawa, *Protein Sci.*, in press (1994).
(32) J.L. Hargrove, H.A. Scoble, W.R. Mathews, B.R. Baumstark and K. Biemann, *J. Biol. Chem.*, **264**, 45 (1989).
(33) P.K. Mehta, T.I. Hale and P. Christen, *Eur. J. Biochem.*, **186**, 249 (1989).
(34) S. Levy and A. Danchin, *Mol. Microbiol.*, **2**, 777 (1988).
(35) P. Bork and K. Rohde, *Biochem. Biophys. Res. Com.*, **171**, 1319 (1990).
(36) P.J. Kraulis, *J. Appl. Cryst.*, **24**, 946 (1991).